A GRAPHICAL USER INTERFACE FOR A PROJECT GLOSSARY GENERATOR Victoria Huang, Marek Gibiec, Paula Laurent, and Jane Cleland-Huang*, School of Computing, DePaul University, Chicago, IL, 60604, jhuang@cs.depaul.edu

A project glossary is utilized in a software development project to document key terminology and its associated definitions. Ideally it is created early in a software development lifecycle and then used consistently within requirements, design, code, and test documents. Use of a project glossary is considered a sound software engineering practice that reduces ambiguity during the initial requirements elicitation and analysis process, and increases understandability and maintainability of the system in general. Unfortunately project stakeholders often fail to create a project glossary or create one but fail to use it consistently across the software development lifecycle. In such cases, it is useful to be able to reconstruct a glossary through analyzing existing project documentation. In prior work we developed a tool for automatically generating a project glossary from a requirements specification. The tool uses a part-of-speech tagger named Q-Tag, to identify nouns and noun phrases which serve as candidate glossary items. The domain specificity of each item is then computed by comparing the frequency at which it occurs in the targeted requirements specification against the probability of its occurrence in a corpus of general documents. Non domain specific items and items that occur only once or twice in the requirements specification are filtered out, and the remaining items are presented to a user as candidate glossary terms.

This particular project involved developing a graphical user interface to serve as a front end to the project glossary generator. The tool allows a user to select a requirements specification document for which a project glossary is to be generated, and to select up to 100 documents to serve as the corpus of domain documents. The user then selects a feature to execute the underlying code in order to generate a list of candidate glossary terms and phrases. These results are then displayed on a second screen, where the user has the opportunity to sort the results or filter them according to upper and lower threshold values for frequency and domain specificity. Finally the user selects a subset of the terms and phrases for inclusion in the project glossary. Our prior studies across several datasets showed that the tool was able to recreate a significant number of glossary terms and also identified an additional set of useful terms that could be added to the glossary and used throughout the remainder of the project.